



Knowledge-based generalization of metabolic networks: a practical study

Anna Zhukova, David James Sherman

► To cite this version:

Anna Zhukova, David James Sherman. Knowledge-based generalization of metabolic networks: a practical study. *Journal of Bioinformatics and Computational Biology*, 2014, 12(2) (1441001), 10.1142/S0219720014410017 . hal-00906911

HAL Id: hal-00906911

<https://inria.hal.science/hal-00906911>

Submitted on 20 Nov 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Journal of Bioinformatics and Computational Biology
 © Imperial College Press

KNOWLEDGE-BASED GENERALIZATION OF METABOLIC NETWORKS: A PRACTICAL STUDY

ANNA ZHUKOVA

*Inria / Université Bordeaux / CNRS joint project-team MAGNOME,
 F-33405 Talence, France
 anna.zhukova@inria.fr*

DAVID J SHERMAN

*Inria / Université Bordeaux / CNRS joint project-team MAGNOME,
 F-33405 Talence, France
 david.sherman@inria.fr*

Received (Day Month Year)

Revised (Day Month Year)

Accepted (Day Month Year)

The complex process of genome-scale metabolic network reconstruction involves semi-automatic reaction inference, analysis, and refinement through curation by human experts. Unfortunately, decisions by experts are hampered by the complexity of the network, which can mask errors in the inferred network. In order to aid an expert in making sense out of the thousands of reactions in the organism's metabolism, we developed a method for knowledge-based generalization that provides a higher-level view of the network, highlighting the particularities and essential structure, while hiding the details.

In this study, we show the application of this generalization method to 1286 metabolic networks of organisms in *Path2Models* that describe fatty acid metabolism. We compare the generalized networks and show that we successfully highlight the aspects that are important for their curation and comparison.

Keywords: Metabolic modeling; generalization; genome-scale reconstruction; fatty acid metabolism.

1. Introduction

In this work we attempt to resolve the apparent conflict between the very detailed description of metabolic networks that are needed for accurate computer simulation, and the higher-level abstract views of the network that are needed by their human curators. Genome-scale metabolic networks are complex systems that describe thousands of reactions participating in the organism's metabolism. Human curators, who generally speaking understand best networks that have hundreds of reactions, are distracted by these details and cannot easily identify the reactions that require their intervention. An appropriate level of abstraction needs to be found to allow experts to explore and compare whole genome networks.

This is even more important as metabolic networks for more and more organisms are being *inferred* and stored in biological network collections, such as the Biocompare database¹ and BiGG². During the process of network reconstruction, their reactions are semi-automatically inferred from pathway and reaction databases such as *KEGG*³ and Reactome⁴, and from existing networks for similar organisms using genomic data⁵. The inferred draft network is then refined during several iterations of error detection, gap filling, analysis and improvement^{5,6}. Examples of network inferring tools include Pathway Tools⁷, IdentiCS⁸ and MetaNetX.org⁹.

Even though automatic tools for network inference and analysis are becoming more powerful, analysis by a human expert remains necessary. A human expert in organism-specific metabolism can find erroneous reactions in the draft network, and add organism-specific ones that could not be inferred automatically. But since genome-scale networks are tailored for computer simulation, and thus include all the reactions thought to participate in the organism's metabolism, these networks are quite complicated and errors may be hidden in the multitude of reactions.

To help a human expert in understanding these detailed networks, we developed a method for knowledge-based generalization (in prep.) that focusses on the higher-level relationships in the network, while omitting the details. The goal of this manuscript is to evaluate this method.

2. Network generalization method

The generalization method operates on networks in *SBML*¹⁰ format. *SBML* elements include biochemical species present in the network, reactions between them, and the compartments in which they are located. *SBML* allows annotation of its elements with ontological terms, e.g. species can be annotated with the *ChEBI* ontology¹¹ identifiers.

The Enzyme Commission number (EC number) is a top-down hierarchical classification scheme for enzymes, based on the chemical reactions they catalyse.¹² In theory, climbing the classification hierarchy up to a higher level of abstraction for the reactions present in the network would suffice to generalize the network. Unfortunately, many of existing networks lack EC number annotations. Consequently, our method groups reactions in the network based on the biochemical species they operate on.

The method first groups the species present in the network into semantically equivalent classes. The appropriate level of abstraction for those classes is defined by the network itself as the most general one that satisfies two restrictions:

- (i) species that participate in the same reaction cannot be grouped together (to preserve the stoichiometry of the reactions in the network),
- (ii) species that do not participate in any pair of similar reactions are not grouped together (as there is no evidence of their similarity in the network).

Possible groups for the species are chosen based on hierarchical relationships in the

ChEBI ontology. For instance, *(S)*-3-hydroxydecanoyl-CoA, *(S)*-3-hydroxyloauroyl-CoA and *(S)*-3-hydroxytetradecanoyl-CoA have a common ancestor *hydroxy fatty acyl-CoA* in *ChEBI*. If there exist similar reactions operating with them in the network, e.g. *3-hydroxyacyl-CoA dehydratase*, and no reaction whose stoichiometry would be broken by such a generalization, then they can be grouped and generalized into *hydroxy fatty acyl-CoA*.

Reactions that share the same generalized reactants and the same generalized products, are considered equivalent and are factored together into a generalized reaction. This process is illustrated in Figure 1.

We do not generalize frequently occurring species, e.g. *oxygen*, *hydrogen*, *water*, *ATP*. Grouping species increases the number of reactions they participate in, while these are already shared by many reactions and networks to such an extent that during visualisation these species are usually duplicated¹³ to improve readability.

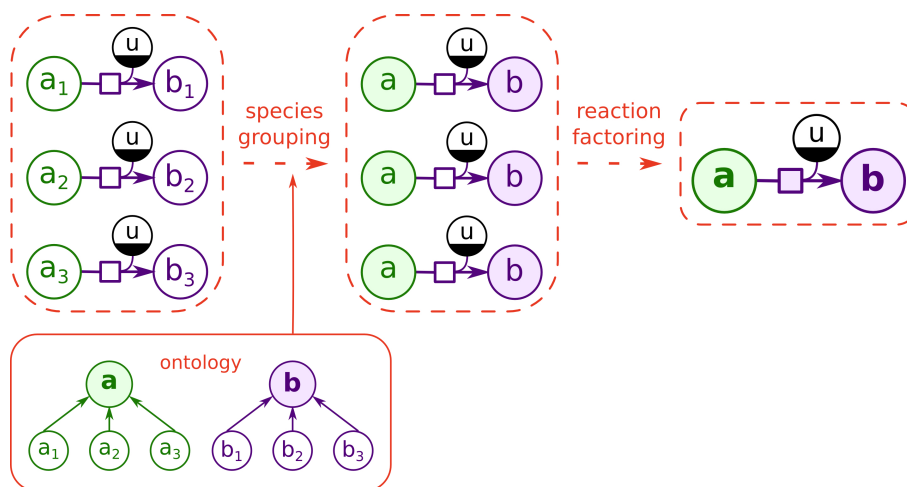


Fig. 1. The generalization method divides the species into ubiquitous (black) and specific, and then generalizes them based on the hierarchical relationships in the *ChEBI* ontology, and reactions they participate in. The reactions between equivalent species are grouped together into a generalized reaction.

3. Applications

In order to demonstrate how the generalization method helps to detect problems and particularities in metabolic networks, we applied it to 1286 metabolic networks that describe the same process in as many different organisms. For our evaluation we chose *fatty acid metabolism*, both because it is a well-studied target for biotechnology applications, and because its presence or absence in different phylogenetic clades is generally known. We downloaded the networks that describe fatty acid

metabolism from *Path2Models* project. *Path2Models* is a branch of the *Biomodels* database, that stores networks that were automatically generated from *KEGG* pathways.

The process of *beta-oxidation of fatty acids*¹⁴ repeats four main steps:

- (i) *dehydration*, transforming *fatty acyl-CoA* into *dehydroacyl-CoA*,
- (ii) *hydration*, transforming *dehydroacyl-CoA* into *hydroxyacyl-CoA*,
- (iii) *oxidation*, transforming *hydroxyacyl-CoA* into *3-ketoacyl-CoA*, and
- (iv) *thiolysis*, transforming *3-ketoacyl-CoA* into *acetyl-CoA* and *fatty acyl-CoA* with a two carbons shortened chain.

A long chain of reactions, repeating these four steps again and again while transforming a long chain *fatty acyl-CoA* into a short chain one, becomes a cycle in a generalized network: the reactions operating with the *fatty acyl-CoA* species of different carbon chain length, corresponding to each of the steps, are factored together into four generalized reactions (see Figure 2).

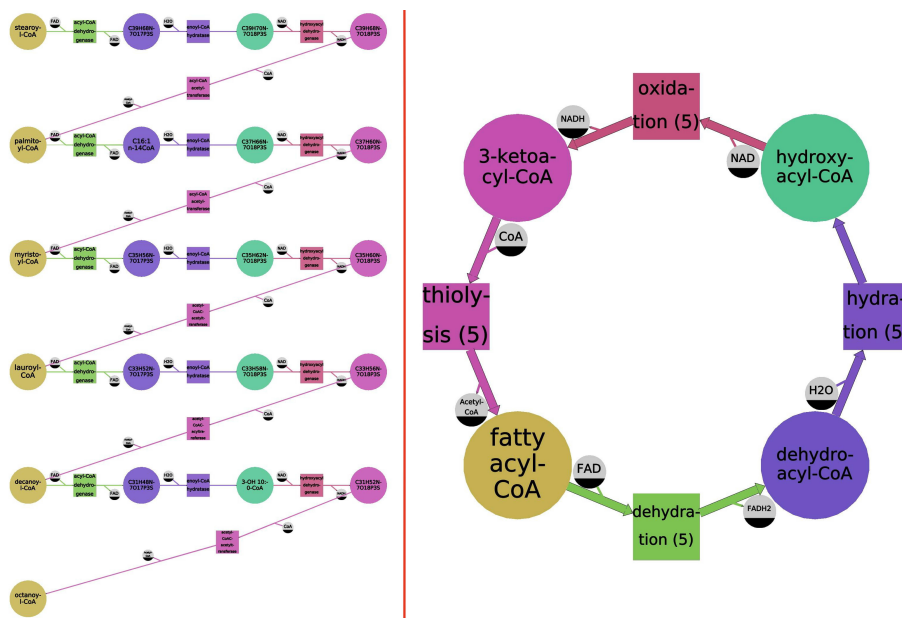


Fig. 2. The initial (left) and generalized (right) representation of *beta-oxidation of fatty acids*. The number in parentheses in each generalized reaction shows how many specific reactions were grouped together.

Among the 1286 networks that have been generalized, 243 do not have the *beta-oxidation* pathway at all, and 124 have the complete *beta-oxidation* cycle present.

3.1. Missing steps

If an enzyme catalysing some of the reactions is missing in the network, then the generalized representation is not a cycle any more. For example, if *EC 1.1.1.35* is missing, the whole group of *oxidation* reactions participating in the network is eliminated, breaking the cycle (see Figure 3). This is more evident on a generalized network than on an initial one, where the absence of these reactions might be hidden by the abundance of other reactions.



Fig. 3. The generalized representation of *beta-oxidation of fatty acids* of *BMID000000136479* (oleaginous yeast *Y. lipolytica*, noncurated network from *Path2Models*). The *oxidation* reaction is missing.

Among the generalized networks, 128 have one step missing, 95 of them miss *oxidation*, 23 lack *dehydration*, 8 do not have *hydration*, and only 2 (*BMID000000046743* and *BMID000000129004*) miss *thiolysis*. As the most of the *beta-oxidation* pathway is present, it is probable that the absence of this step is an error in the reconstruction process. For example, in network *BMID000000136479*, which represents *fatty acid metabolism* in the yeast *Yarrowia lipolytica* (strain *CLIB 122/E 150*), the *oxidation* step is missing (Figure 3); while in the generalized network of the curated genome-scale network of the same strain of the same organism *MODEL1111190000*¹⁵, the *beta-oxidation* cycle is complete (Figure 4). By helping to draw the curator's attention to such missing steps, generalization can improve the speed and accuracy of network curation. Generalization can highlight missing steps by showing broken cycles, but also by showing changes in the *path profile*, the number of grouped reactions along a path (Figure 3).

On the other hand 145 networks have two steps missing, and 646 have only one out of the four generalized reactions present.

3.2. Alternative steps

In addition to missing steps, the generalization of the network can highlight alternative paths that may be shortcuts or represent substrate specificities. It is important that such paths not be hidden in the generalized network, as they are often the cases that require the human expert decision as to whether these alternatives appear due to an error or to an organism-specific adaptation.

In the case of *beta-oxidation*, an example of reaction variations are two versions of the *oxidation* reaction that use different ubiquitous species, as shown in Figure 5. Among the networks that we analysed, it is the only reaction that may have

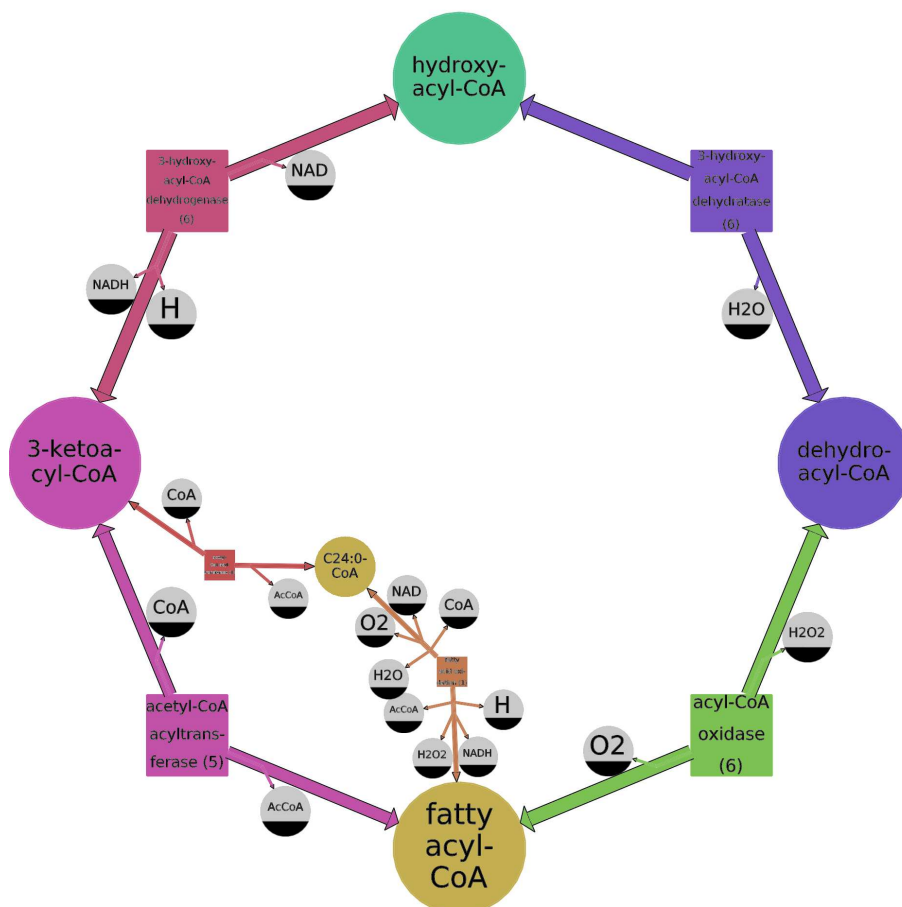


Fig. 4. The generalized representation of *beta-oxidation of fatty acids* of *MODEL1111190000* (*Y. lipolytica*, curated network from Loira et al. 2012). The cycle is complete.

variations within the same network, indeed, 168 out of 170 networks that have the *oxidation* reaction present, have it in two versions.

Complete statistics on missing and alternative *beta-oxidation* steps in the analysed networks are shown in Table 1. Changes in the numbers of grouped reactions in a profile path can also be used to evaluate alternative paths (data not shown).

4. Comparing generalized networks

By abstracting detailed networks, generalization makes it easier to compare them at different scales of divergence. Since each generalization is maximal, as determined by the actual reactions and biochemical species in the network, it masks unimportant differences in the intermediate levels of the ontologies of chemical entities and

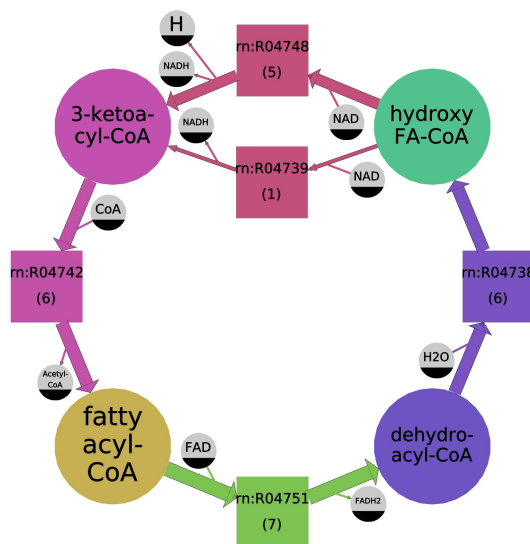


Fig. 5. The generalized representation of beta-oxidation of fatty acids of *BMID000000103487* (nonpathogenic bacterium *Burkholderia thailandensis*). Two variants of the *oxidation* reaction (top) are present.

reactions. Restrictions *i* and *ii* (section 2) guarantee that any differences between two networks that remain after generalization result from real differences in their network structure. Furthermore, generalization makes these differences stand out from the structure of the conserved generalized network.

For example, comparing the standard *beta-oxidation* pathway (Figure 2) and those for *Y. lipolytica* (Figure 4) and *B. thailandensis* (Figure 5) very clearly shows the specificities of the two latter networks, as well as the biochemical species that prevent generalization. In *Y. lipolytica* C24:0-CoA is specially handled by specific acyl-transferase and fatty acid oxidation enzymes; in *B. thailandensis* a specific dehydrogenase is used for oxidation in one case out of six.

To explore the effect of network generalization on a broad evolutionary scale, we first mapped the 1286 networks to the NCBI taxonomy database¹⁶ and compared *beta-oxidation* pathway configurations between superkingdoms (Tables 1 and 2). The analysed networks represent *fatty acid metabolism* in 138 *eukaryota*, 1045 *bacteria* and 103 *archaea* species.

The percentage of species for which the four-step *beta-oxidation of fatty acids* is not present, or only one out of the four reactions is available (thus most probably used in a different pathway) is similar (about 60%) for all the superkingdoms. The case when the complete cycle is present diverges more. The complete cycle appears in some *eukaryota* and *bacteria*, but not in any of the 103 analysed *archaea* networks. This situation is supported by the MetaCyc pathway database¹⁷: The

beta-oxidation pathway is present for *eukaryota* and *bacteria*, but not for *archaea*. This might be explained by the fact that in spite of the presence in most *archaea* of the gene candidates for degradation of activated *fatty acids* via the *beta-oxidation* pathway, *archaea* do not encode components of a *fatty acid synthase* complex.¹⁸

Table 1. Presence of reactions of the *generalized beta-oxidation of fatty acids* cycle in different networks **across the three superkingdoms** (● stands for two versions of the corresponding reaction present in the network).

<i>dehyd- ration</i>	<i>hyd- ration</i>	<i>oxi- dation</i>	<i>thio- lysis</i>	all networks	number of		
					eukaryota	bacteria	archaea
●	●	●●	●	124	4	120	
●	●	-	●	95	33	44	18
-	●	●●	●	23		23	
●	-	●●	●	8		8	
●	●	●●	-	2	1	1	
●	-	-	●	68	14	48	6
-	●	-	●	63		44	19
-	-	●●	●	10		10	
●	●	-	-	2	2		
-	●	●●	-	1	1		
●	-	●	-	1		1	
●	-	-	-	430	65	365	
-	-	-	●	166	13	93	60
-	●	-	-	49		49	
-	-	●	-	1	1		
-	-	-	-	243	4	239	
<i>Total:</i>				1286	138	1045	103

Table 2. Percentage of different *generalized beta-oxidation of fatty acids* cycle configurations in different networks.

<i>beta-oxidation cycle configuration</i>	all networks	% of		
		eukaryota	bacteria	archaea
complete cycle	10%	3%	11%	0%
one step missing	10%	25%	7%	18%
two steps missing	11%	12%	10%	24%
three steps missing	50%	57%	49%	58%
all steps missing	19%	3%	23%	0%

To further explore how generalization can help compare networks across evolu-

tionary ranges, we considered the *beta-oxidation* pathway in 47 fungal species (Table 3). The first striking result is that the KEGG pathway method used by *Path2Model* seems to systematically miss the oxidation enzyme (column 3), since it is absent for almost all fungal networks yet fatty acid metabolism is a very common pathway. The second is that dehydration and thiolysis enzymes (columns 1 and 4) are almost always present, which is surprising, but since these are large classes of enzymes that are present in other pathways, perhaps many of the enzymes in these columns are misassigned to this pathway. What remain are the hydration enzymes (column 2), which show some variation between the networks in Table 3. In many cases these enzymes are absent in known pathogens, such as the *Candida*, which hints that these species may obtain the fatty acids from the host rather than through synthesis. However, the systematic biases seen in the other columns make it impossible to find the correlated gene losses that are the hallmark of missing pathways.

Significantly, this shows that network generalization is an excellent tool for abstracting networks from very different lifestyles up to a comparable level of complexity, that directly reveals species-specific differences and systematic biases. These are precisely the clues that human curators would need in order to judge to what degree the *beta oxidation* pathway is present in each of these species.

5. Discussions

Using our network generalization method we have studied 1286 networks describing metabolism of fatty acids in as many organisms. Generalization helps a human to understand, compare and classify those networks. Providing a higher-level view of the network by factoring the abundance of similar reactions, it allows for easier comprehension of the general network structure, and highlights possible problems and organism-specific particularities. Generalization highlights potential errors in inferred draft networks, exposes specific absences or alternatives, and makes it possible to compare networks between species, clades, and kingdoms at a higher level of abstraction.

Generalization can also help in finding a standard template for a pathway, using which curators can analyse this pathway in the organism of interest. In our example, the complete *beta-oxidation* cycle without alternative steps, served as a standard template.

We have studied the correlation between differences in the generalized networks, and their belonging to one of the three superkingdoms: *eukaryota*, *bacteria* or *archaea*. Our method highlighted known tendencies of these superkingdoms, such as the absence of *beta-oxidation* in *archaea*. However, a far more interesting goal is to understand the differences between networks of closely related organisms, in order to study the connection between the differences in generalized metabolic properties of organisms and the differences in their physiology in more refined details, comparing phylogenetically close species, or even different strains of the same organism. Network generalization will expose the absent reactions or the alternative paths

that characterize individual species or strains, and more closely establish the link between genotype and phenotype.

Acknowledgements

The authors would like to thank Dr. Pascal Durrens for valuable comments of fatty acid metabolism. We also thank Dr. Romain Bourqui and Dr. Antoine Lambert for advice on graph layout.

Anna Zhukova is supported by an Inria CORDI-S doctoral fellowship.

References

1. Li C, Donizelli M, Rodriguez N, Dharuri H, Endler L, Chelliah V, Li L, He E, Henry A, Stefan MI, Snoep JL, Hucka M, Le Novère N, Laibe C, BioModels Database: An enhanced, curated and annotated resource for published quantitative kinetic models. *BMC Systems Biology* 2010, **4**:92.
2. Schellenberger J, Park JO, Conrad TM, Palsson BO, BiGG: a Biochemical Genetic and Genomic knowledgebase of large scale metabolic reconstructions. *BMC bioinformatics* 2010, **11**:213.
3. Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M, KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res.* 2012, **40**(0305-1048 (Linking)):D109–14.
4. Milacic M, Haw R, Rothfels K, Wu G, Croft D, Hermjakob H, D'Eustachio P, Stein L, Annotating Cancer Variants and Anti-Cancer Therapeutics in Reactome. *Cancers* 2012, **4**(4):1180–1211.
5. Thiele I, Palsson BO, A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nature protocols* 2010, **5**:93–121.
6. Swainston N, Smallbone K, Mendes P, Kell D, Paton N, The SuBliMinaL Toolbox: automating steps in the reconstruction of metabolic networks. *Journal of integrative bioinformatics* 2011, **8**(2):186.
7. Karp PD, Paley S, Romero P, The Pathway Tools software. *Bioinformatics* 2002, **18**(Suppl 1):S225–S232.
8. Sun J, Zeng AP, IdentiCS - Identification of coding sequence and in silico reconstruction of the metabolic network directly from unannotated low-coverage bacterial genome sequence. *BMC Bioinformatics* 2004, **5**:112.
9. Ganter M, Bernard T, Moretti S, Stelling J, Pagni M, MetaNetX.org: a website and repository for accessing, analysing and manipulating metabolic networks. *Bioinformatics (Oxford, England)* 2013, **29**(6):815–6.
10. Hucka M, Hoops S, Keating SM, Le Novère N, Sahle S, Wilkinson DJ, Systems Biology Markup Language (SBML) Level 2: Structures and Facilities for Model Definitions 2008.
11. de Matos P, Alcántara R, Dekker A, Ennis M, Hastings J, Haug K, Spiteri I, Turner S, Steinbeck C, Chemical Entities of Biological Interest: an update. *Nucleic Acids Research* 2010, **38**(suppl 1):D249–D254.
12. Fleischmann A, Darsow M, Degtyarenko K, Fleischmann W, Boyce S, Axelsen KB, Bairoch A, Schomburg D, Tipton KF, Apweiler R, IntEnz, the integrated relational enzyme database. *Nucleic acids research* 2004, **32**(Database issue):D434–7.
13. Rohn H, Junker A, Hartmann A, Grafahrend-Belau E, Treutler H, Klapperstück M, Czauderna T, Klukas C, Schreiber F, VANTED v2: a framework for systems biology applications. *BMC systems biology* 2012, **6**:139.
14. Metzler DE, *Biochemistry: The Chemical Reactions of Living Cells*. No. v. 1 in Biochemistry: The Chemical Reactions of Living Cells, Elsevier Science 2001.
15. Loira N, Dulermo T, Nicaud JM, Sherman D, A genome-scale metabolic model of the

- lipid-accumulating yeast *Yarrowia lipolytica*. *BMC Systems Biology* 2012, **6**:35.
16. Sayers EW, Barrett T, Benson DA, Bryant SH, Canese K, Chetvernin V, Church DM, DiCuccio M, Edgar R, Federhen S, Feolo M, Geer LY, Helmberg W, Kapustin Y, Landsman D, Lipman DJ, Madden TL, Maglott DR, Miller V, Mizrachi I, Ostell J, Pruitt KD, Schuler GD, Sequeira E, Sherry ST, Shumway M, Sirotkin K, Souvorov A, Starchenko G, Tatusova TA, Wagner L, Yaschenko E, Ye J, Database resources of the National Center for Biotechnology Information. *Nucleic acids research* 2009, **37**(Database issue):D5–15.
 17. Caspi R, Altman T, Dreher K, Fulcher CA, Subhraveti P, Keseler IM, Kothari A, Krummenacker M, Latendresse M, Mueller LA, Ong Q, Paley S, Pujar A, Shearer AG, Travers M, Weerasinghe D, Zhang P, Karp PD, The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic acids research* 2012, **40**(Database issue):D742–53.
 18. Falb M, Müller K, Königsmaier L, Oberwinkler T, Horn P, von Gronau S, Gonzalez O, Pfeiffer F, Bornberg-Bauer E, Oesterhelt D, Metabolism of halophilic archaea. *Extremophiles : life under extreme conditions* 2008, **12**(2):177–96.



Anna Zhukova received her Specialist degree in Mathematics and Software Engineering from St. Petersburg State University, Russia in 2010. In 2011 she started her PhD research in MAGNOME project-team in Inria Bordeaux Sud-Ouest research center, France. Her research is focused on knowledge-based scaling for biological networks.



David James Sherman (University of Chicago S.B.'86, S.M.'89, Ph.D.'94) is a Senior Scientist with Inria and leader of the MAGNOME project-team. Originally interested in formal methods for industrial systems, he turned in 1999 to modeling complex biological systems and large-scale comparative genomics. His interests include knowledge mining and classification in bioinformatics, inference of models and networks, algorithms and strategies for genome comparison; and overuse of hendiadys.

Table 3. Presence of reactions of the *generalized beta-oxidation of fatty acids* cycle in different networks of **fungi** genomes.

<i>fungi</i> species	<i>dehyd-</i> <i>ration</i>	<i>hyd-</i> <i>ration</i>	<i>oxi-</i> <i>dation</i>	<i>thio-</i> <i>lysis</i>
Microsporidia				
. Encephalitozoon cuniculi GB-M1	-	-	-	•
Dikarya/Ascomycota				
. Taphrinomycotina				
. . Schizosaccharomyces pombe 972h-	-	-	-	•
. saccharomyceta				
. . Saccharomycotina/Saccharomycetales				
. . . Metschnikowiaceae				
. . . . Clavispora lusitaniae ATCC 42720	•	-	-	•
. . . Debaryomycetaceae				
. . . . Lodderomyces elongisporus NRRL YB-4239	•	-	-	•
. . . . Scheffersomyces stipitis CBS 6054	•	•	-	•
. . . . Meyerozyma guilliermondii ATCC 6260	•	-	-	•
. . . . Debaryomyces hansenii CBS767	•	•	-	•
. . . Dipodascaceae				
. . . . Yarrowia lipolytica CLIB122	•	•	-	•
. . . Saccharomycetaceae				
. . . . Komagataella pastoris GS115	•	-	-	•
. . . . Zygosaccharomyces rouxii CBS 732	•	-	-	•
. . . . Lachancea thermotolerans CBS 6340	•	-	-	•
. . . . Saccharomyces ceremonial S288c	•	-	-	•
. . . . Vanderwaltozyma polyspora DSM 70294	•	-	-	•
. . . . Ashbya gossypii ATCC 10895	•	-	-	•
. . . . Candida glabrata CBS 138	•	•	-	•
. . . . Kluyveromyces lactis NRRL Y-1140	•	•	-	•
. . . mitosporic Saccharomycetales/Candida				
. . . . Candida dubliniensis CD36	•	-	-	•
. . . . Candida tropicalis MYA-3404	•	-	-	•
. . . . Candida albicans SC5314	•	-	-	•
. . Pezizomycotina				
. . . Pezizomycetes				
. . . . Tuber melanosporum Mel28	•	•	-	•
. . . leotiomyceta				
. . . . dothideomyceta/Phaeosphaeria nodorum SN15	•	•	-	•
. . . . sordariomyceta				
. Leotiomycetes/Sclerotiniaceae				
. Sclerotinia sclerotiorum 1980 UF-70	•	•	-	•
. Botryotinia fuckeliana B05.10	•	•	-	•

Table 3 (*Continued*). Presence of reactions of the *generalized beta-oxidation of fatty acids* cycle in different networks of **fungal genomes**.

<i>fungal species</i>	<i>dehyd- ration</i>	<i>hyd- ration</i>	<i>oxi- dation</i>	<i>thio- lysis</i>
. Sordariomycetes				
. Sordariomycetidae				
. Podospora anserina S mat+	•	•	-	•
. Neurospora crassa OR74A	•	•	-	•
. Magnaporthe oryzae 70-15	•	•	-	•
. Hypocreomycetidae				
. Fusarium graminearum PH-1	•	-	-	-
. . . . Eurotiomycetes/Eurotiomycetidae				
. . . . Eurotiales/Aspergillaceae				
. Penicillium chrysogenum Wisconsin 54-1255	•	-	-	-
. Neosartorya fischeri NRRL 181	•	•	-	•
. Aspergillus oryzae RIB40	•	-	-	-
. Aspergillus niger CBS 513.88	•	-	-	-
. Aspergillus clavatus NRRL 1	•	•	-	•
. Aspergillus flavus NRRL3357	•	-	-	-
. Aspergillus fumigatus Af293	•	•	-	•
. Aspergillus nidulans FGSC A4	•	-	-	-
. . . . Onygenales				
. Uncinocarpus reesii 1704	•	•	-	•
. Coccidioides immitis RS	•	•	-	•
. Coccidioides posadasii C735 delta SOWgp	•	•	-	•
Dikarya/Basidiomycota				
. Ustilaginomycotina				
. . Malassezia globosa CBS 7966	•	-	-	-
. . Ustilago maydis 521	•	-	-	-
. Agaricomycotina				
. . Agaricomycetes				
. . . Postia placenta Mad-698-R	•	•	-	-
. . . Schizophyllum commune H4-8	•	•	-	•
. . . Moniliophthora perniciosa FA553	•	•	-	•
. . . Laccaria bicolor S238N-H82	•	•	-	•
. . . Coprinopsis cinerea okayama7#130	•	•	-	•
. . Tremellomycetes/Cryptococcus neoformans				
. . . Cryptococcus neoformans var. neoformans B-3501A	•	-	-	-
. . . Cryptococcus neoformans var. neoformans JEC21	•	-	-	-